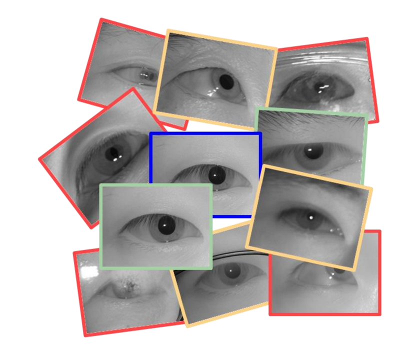
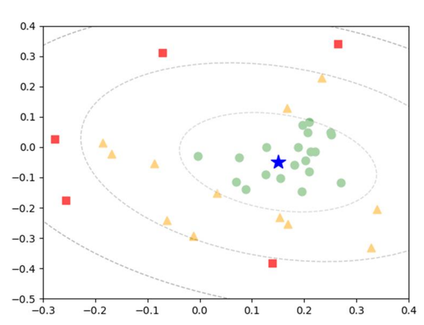
Dual Dense Uncertainty Embedding for Iris Recognition

Yunlong Wang[1], Jianze Wei[2], Zhaofeng He[3], Zhenan Sun[1], Tieniu Tan[1,4]
[1]Center for Research on Intelligent Perception and Computing (CRIPAC), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, P.R.China
[2]Institute of Microelectronics, Chinese Academy of Sciences
[3]School of Artificial Intelligence, Beijing University of Posts and Telecommunications
[4]Nanjing University, Nanjing, P.R.China

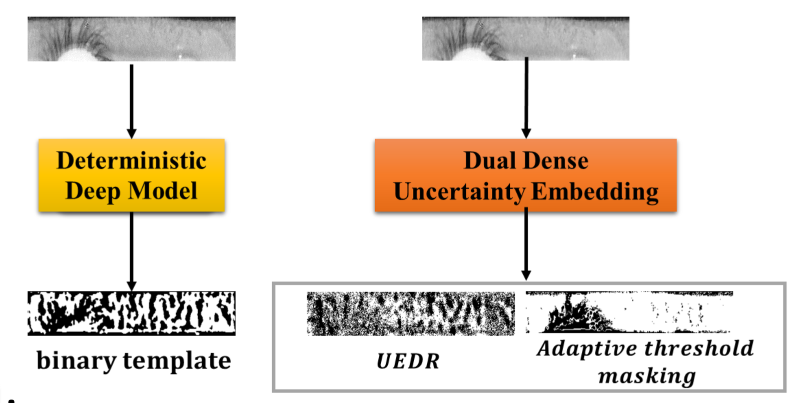
IEEE International Joint Conference on Biometrics (IJCB 2023)
http://www.cripacsir.cn/

Motivation


Captured Iris Images of One Person


Embeddings in Feature Space

Uncertain acquisition factors inevitably affect the process of iris imagery formation. Deep uncertainty embedding (DUL) is leveraged to represent the iris image using a Gaussian distribution.



The binary mask indicating valid iris regions is solely determined by a fixed threshold or the output of standalone segmentation and localization algorithms. Deterministic threshold masking strategy is suboptimal considering the versatile uncertainties in the iris images collected from various scenarios.

Framework

- **Uncertainty Embedded Dense Representation (UEDR)**

Model each pixel's uncertainty by a univariate Gaussian distribution, μ-branch and σ-branch are constructed by a simple transformation with a convolutional layer and an instance normalization (IN) layer

$$\mathcal{P}_\mu = IN_\mu(Conv_\mu(\mathcal{H}))$$

$$\mathcal{P}_\sigma = IN_\sigma(Conv_\sigma(\mathcal{H}))$$

Training phase

$$\mathcal{S}_{ij} = \mathcal{P}_\mu^{ij} + \varepsilon_{ij} \times \mathcal{P}_\sigma^{ij}$$

Inference phase

$$\mathcal{S}_{ij} = \mathcal{P}_\mu^{ij}$$

- **Adaptive Threshold Masking**

Find a optimal α for each pixel in an iris image through predicting a variance scaling map (VSM), taking not only the intensity distribution of the iris image but also each pixel's low-level uncertainty into consideration.

$$\mathcal{A} = Sigmoid(Conv(\mathcal{H}))$$

$$\mathcal{M}_{ij} = \begin{cases} 1, & if\ \left|I_{iris}^{ij} - I_\mu\right| < 3\mathcal{A}_{ij} I_\sigma \\ 0, & otherwise \end{cases}$$

- **Loss function**

Extended Triplet Loss (ETL)

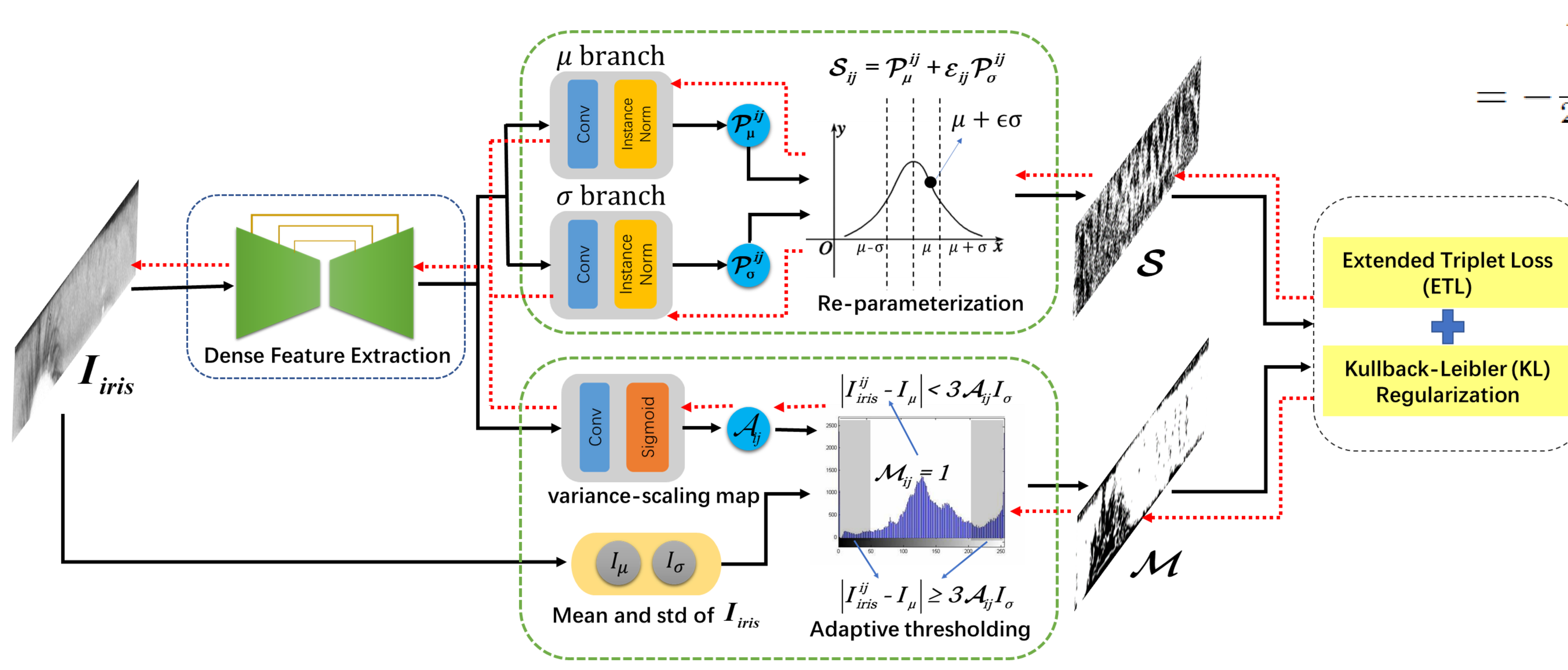
$$\mathcal{L}_{etl} = \frac{1}{M}\sum_{i=1}^{M} \max(D(\mathcal{S}_i^P, \mathcal{S}_i^A) - D(\mathcal{S}_i^N, \mathcal{S}_i^A) + \tau, 0)$$

Kullback-Leibler (KL) Regularization

$$\mathcal{L}_{kl} = \mathrm{E}\left\{KL\left[\mathcal{N}(\mathcal{S}_{ij}|\mu_{ij}, \sigma_{ij}) \parallel \mathcal{N}(\varepsilon_{ij}|0,1)\right]\right\}$$
$$= -\frac{1}{2NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}(1 + \log\sigma_{n,i,j}^2 - \mu_{n,i,j}^2 - \sigma_{n,i,j}^2)$$

Total loss

$$\mathcal{L}_{total} = \mathcal{L}_{etl} + \lambda\mathcal{L}_{kl}$$

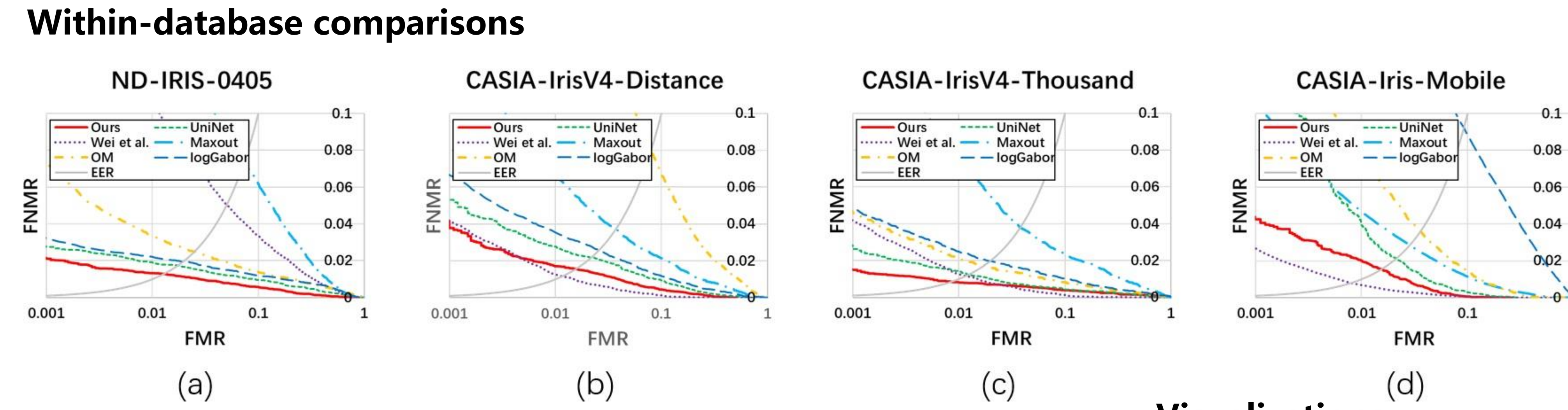


The schematic of the proposed framework ***D²UE***

Experimental Results

Within-database comparisons

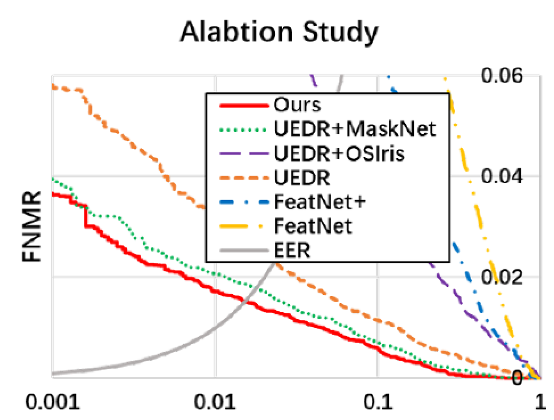
		logGobar	Ordinal	Maxout	Wei et al.	UniNet†	Ours
ND0405	EER	1.94	2.49	7.33	5.05	1.74	**1.61** (↓0.13)
	FNMR	3.24	7.51	32.31	20.90	2.75	**2.11** (↓0.64)
Distance	EER	2.60	8.01	3.73	**1.17**	2.17	1.57 (↓0.60)
	FNMR	6.61	40.73	14.85	4.18	5.32	**3.95** (↓1.37)
Thousand	EER	1.97	1.75	3.78	1.17	1.22	**0.85** (↓0.37)
	FNMR	4.81	4.62	16.24	4.17	2.73	**1.51** (↓1.22)
Mobile	EER	9.24	3.62	2.85	**0.79**	2.11	1.56 (↓0.55)
	FNMR	37.94	19.73	10.58	**2.67**	12.50	4.25 (↓8.25)



(a) (b) (c) (d)

Ablation study

	No Iris Mask	Deterministic Threshold Masking	Adaptive Threshold Masking	EER
FeatNet	✔			11.97
FeatNet+	✔			7.61
UEDR	✔			2.27
UEDR+*OSIris*		✔		5.25
UEDR+*MaskNet*		✔		1.79
Ours			✔	**1.57**



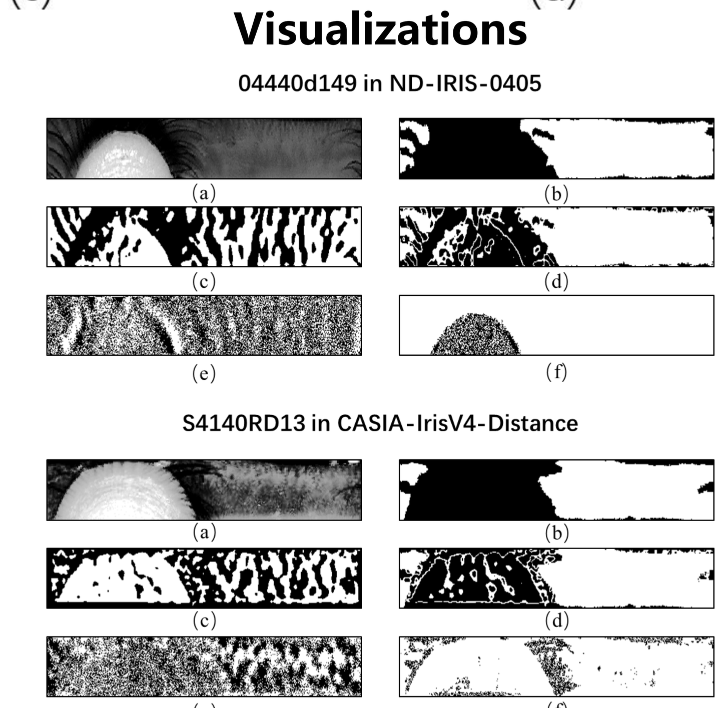
Cross-database comparisons

	UniNet		Ours	
	EER	FNMR	EER	FNMR
Distance	7.55	23.48	7.04	23.21
Thousand	3.10	10.68	5.3	13.82
Mobile	1.31	4.87	1.18	2.79

Model complexity

		#Params(K)	MACs(M)
Wei et al.		6558.11	1261.37
UniNet	*FeatNet*	27.46	144.90
	MaskNet	97.86	98.86
	Total	125.32	243.76
Ours		**29.48**	**211.22**

Visualizations



Conclusions

In this paper, D²UE is proposed to model data uncertainty of iris recognition in a pixel-level manner, which can attenuate the interference caused by uncertain acquisition factors. D²UE takes the intermediate feature map of any dense DL framework as input and generates the UEDR and VSM of an iris image. VSM is then leveraged to produce a binary mask through adaptive threshold masking, and thus pixelwise iris segmentation is on longer needed. Experimental results on several public iris datasets demonstrate the superiority of D²UE in improving the recognition performance of baseline methods, and it is a remarkably lightweight building blocking.